Stanley W. Nash

PROFESSOR EMERITUS

B.A. (1939), College of Puget Sound

M.A. (1946), University of California, Berkeley

Ph.D. (1950), University of California, Berkeley

Visiting Associate Professor, Iowa State University (1960-61)

The University of British Columbia (1950-present)

Research Interests

In the past I have worked on various problems in the design of experiments and multivariate analysis. For instance, an experiment may test many genetic strains of some plant and find significant differences. If one then compares a few of the strains which did the very best with those that were the very worst in the first experiment, one may occasionally find no significant differences in the second experiment. I did a thorough analytical study of the power functions of the overall tests of differences in the two experiments and showed how and why the seemingly paradoxical outcomes in the two experiments can frequently occur.

Canonical correlation analysis can be used to carry out a discriminant type of analysis. With a new standardization of the canonical variates, Mahalanobis distances can be calculated between populations using either of the two sets of (random) variables of canonical correlation analysis. This type of analysis was applied to an extensive provenance study of Sitka spruce in British Columbia and Alaska for which both botanical and geographical sets of variables were available.

Similar analyses can be carried out when one considers the correspondence type of analysis of two-way contingency tables. I have done extensive work on this, but never published my results.

At the present time I am interested in problems dealing with three or more sets of (random) variables, which one wants to deal with in a canonical way to accomplish purposes similar to those dealt with by the canonical correlation analysis of two sets of random variables. There have been a number of models proposed, especially in the psychometric literature, but none are entirely satisfactory, particularly when applied to other fields. I would also like to see how such three-mode and higher methods can be adapted to dealing with three-way and higher dimensional contingency tables.

Selected Publications

Nash, S. (1954). "An extension of the Borel-Cantelli lemma." Annals of Mathematical Statistics, **25**, 165-167.

Nash, S. (1956). "Contribution to the theory of experiments with many treatments." University of California Publications in Statistics, **2**, 167-183.

Nash, S. and Ahuja, J.C. (1967). "The generalized Gompertz-Verhulst family of distributions." Sankhya, **A29**, 141-156.

Nash, S. and Falkenhagen, E.R. (1978). "Multivariate classification in provenance research: a comparison of two statistical techniques." Silvae Genetica, **27**, 14-23. Corrections [of misprints], Ibidem, 28, 117.
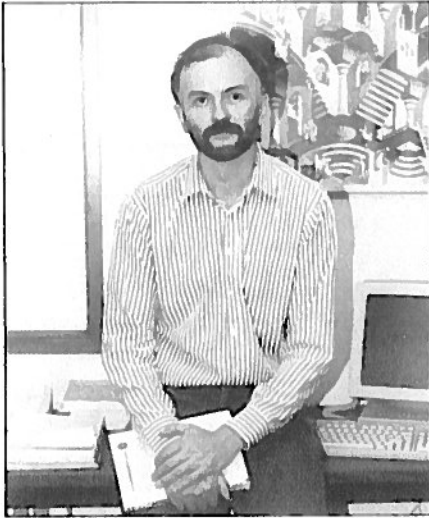
Andrew J. Coldman

ADJUNCT PROFESSOR

B.Sc (1974), University of Sussex,

U.K.M.A. (1975), University of Western Ontario

Dip. Ep. (1977), McGill University

Ph.D. (1987), University of British Columbia

Associate Member, Department of Health Care and Epidemiology, UBC (1990-present)

Statistician, British Columbia Cancer Agency (1981-85)

Head, Biometry Section, British Columbia Cancer Agency (1985-present)

The University of British Columbia (1988-present)

Research Interests

My main research interests are in the area of the application of statistical and probabilistic models to Oncology. My permanent appointment is with the British Columbia Cancer Agency in the division of Epidemiology which conducts a variety of studies into the causes and treatment of cancer and its precursors.

My main research interest of the past several years has been the investigation of the application of branching process models to problems in cancer. Populations of cells, and in particular cancer cells, divide and grow by a process of binary fission. As these populations grow they tend to diversify with cells acquiring characteristics which distinguish them from their fellows. Many such changes are benign and have little significance, however others are critical and affect the curability of the disease. Two examples are the acquisition of drug resistance by cancer cells and the conversion of normal cells into cancer cells. These process may be modelled by branching processes in which the likelihood of an alteration is assumed to be constant at each division. In simple cases recursive algorithms can be developed for the distribution of the number of altered cells. In more complex situations moment generating functions frequently can be found. At present my work in this area has concentrated on the application of these models to the treatment of clinical cancer. Algorithms have been developed for the estimation of parameters from clinical data and the generation of optimal treatment strategies under various constraints.

My other main research interest is in the analysis of epidemiologic data, and in particular data relating to cervical data. A common problem in evaluating screening programs for chronic diseases (such as cervical cancer) is that the introduction of screening biases many of the commonly used measures of disease outcome, e.g., survival. Because of this the need for randomized clinical trials is critical but to date none have been undertaken in pap smear screening for cervical cancer. In the absence of such a trial, analysis of routinely collected data on women undergoing pap smears becomes very important. British Columbia has the oldest screening program in the world and we are currently undertaking analyses, using models of the type proposed by Walter and Day, to examine the effect of different screening frequencies on outcome. Modelling of screening for chronic diseases assumes the existence of a pre-clinical disease state with unknown distribution. The statistical problem is to estimate this distribution and use it to make predictions about optimal screening strategies. These modelling procedures have much in common with those currently in use for the development of AIDS following HIV infection.

Selected Publications

Coldman, A.J., Goldie, J.H. and Ng, V. (1985). "The effect of cellular differentiation on the development of permanent drug resistance." Mathematical Biosciences, 74, 177-198.

Coldman, A.J. and Goldie, J.H. (1986). "Factors affecting the development of permanent drug resistance and their impact upon neo-adjuvant chemotherapy." In: Preoperative (Neoadjuvant) Chemotherapy, Ragaz J., Band, P.R. and J.H. Goldie, (eds.). Springer-Verlag, Berlin, 97-78.

Coldman, A.J., Coppin, C. and Goldie, J.H. (1988). "Models for dose intensity." Mathematical Biosciences, 92, 97-113.

Karsai, M., Coldman, A.J., Gavin, D., Brumelle, S., Boyes, D., Anderson, G. and Benedet, J. (1988). "Cervical intra-epithelial neoplasia in female prisoners in British Columbia." Canadian Medical Journal, 139, 733-736.

Constance van Eeden

ADJUNCT PROFESSOR

B.Sc. (1949), University of Amsterdam

M.Sc. (1954), University of Amsterdam

Ph.D. (1958), University of Amsterdam

Research Associate, Mathematical Center Amsterdam (1954-60)

Visiting Associate Professor, Michigan State University (1960-61)

Research Associate, University of Minnesota (1961-64)

Associate Professor, University of Minnesota (1964-65)

Professeur agrégé, (1965-68) et Professeur titulaire, Université de Montréal, (1968-88)

Visiting Professor, University of Wisconsin, (1969)

Professeur invité, Université de Rennes, (1970)

Chercheure associée, (1989-91) et Professeure associée, Université du Québec à Montréal (1991-present)

Adjunct Professor, The University of British Columbia (1989-present)

Research Interests

My main research interests are in the areas of estimation in restricted parameter spaces, group-Bayes estimation, nonparametrics and selection procedures.

In the problems of estimation in restricted parameter spaces one knows that the parameter to be estimated lies in a known subset of its "natural domain". One might, for example, know that the unknown mean of a normal distribution is nonnegative. In such cases, classical methods like maximum likelihood lead to estimators which are generally inadmissible for squared error loss. I am, and have been, looking for "good" estimators for this kind of problem. As an example, very little is known about minimax estimation in restricted parameter spaces, particularly for the case where the parameter space is not compact.

In group-Bayes estimation, a project on which I am working with Jim Zidek, one of the problems we are looking at is the case where the hyperparameter space is restricted. The search for group-admissible and group-minimax estimators then often leads to questions of admissibility and minimaxity in classical (non-Bayesian) restricted parameter spaces.

In nonparametrics I am working, with Sorana Froda from UQAM, on a saddlepoint approximation to the null distribution of the two-sample Wilcoxon statistic. We obtain the exact coefficients of N^{-i}, $i=0, 1, 2$ and the rate of convergence of the remainder terms. This rate is $O(N^{-3/2})$ for the two-term approximation. This approximation should be an improvement over the normal, the Edgeworth and the "sum-of-independent-uniforms" approximations, particularly in the tails of the distribution. A planned numerical study should tell us in which cases which approximation is best.

In selection procedures problems, on which I am working with Paul van der Laan from the Eindhoven University of Technology, we are looking at a subset selection procedure which uses a continuous loss function instead of the usual 0-1 loss function. For the location problem our loss is zero if the "best" population (that is, the one with largest location parameter) is in the selected subset and an increasing function of the difference (in parameter and value) between the best population and the best one in the selected subset, if not. We are also looking at what happens if the search for the best population is replaced by a search for an ε-best one, that is, a population with a parameter value with ε of the largest parameter value, where ε is a given positive constant.

Selected Publications

van Eeden, C. and Charras, A. (1991) "Bayes and admissibility properties of estimators in truncated parameters spaces." *Canadian Journal of Statistics*, **19**, 121-134.

van Eeden, C. and Charras, A. (1992). "Inadmissibility for squared error loss when the parameter to be estimated is restricted to the interval [a,∞)." Technical Report 114, Department of Statistics, University of British Columbia.

van Eeden, C. and Zidek, J.V. (1993). "Group-Bayes estimation of the exponential mean: A preposterior analysis." Technical Report 120, Department of Statistics, University of British Columbia.

van Eeden, C. and Leslie, J. (1993). "On a characterization of the exponential distribution based on a type 2 right censored sample." *Annals of Statistics*, **21**. To appear.

van Eeden, C. and Zidek, J.V. (1993) "Group Bayes estimation of the exponential mean: a retrospective view of the Wald theory." In: *Proceedings of the Fifth Purdue Symposium on Statistical Decision Theory and Related Topics*. S.S. Gupta and J.O. Berger (eds.). To appear.
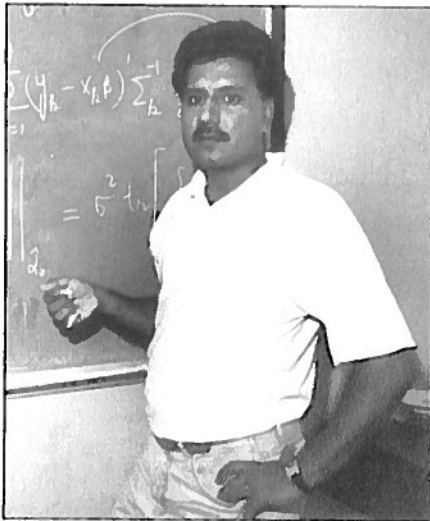
P.S. Gill

HONORARY RESEARCH ASSOCIATE

B.Sc. (1978), Punjab University, Chandigash

M.Sc. (1980), Punjab Agricultural University, Ludhiana

Ph.D. (1987), Indian Institute of Technology, Kanpur

Assistant Professor, Punjabi University, Patiala (1987-88)

Post-Doctoral Fellow, McMaster University (1988-89)

Assistant Professor, Northern Illinois University (1989-90)

Assistant Professor, Memorial University of Newfoundland (1990-92)

College Professor, Okanagan University College (1992-present)

The University of British Columbia (1993-present)

Research Interests

My main field of research is applied statistics in general and experimental design in particular. I have been working on the optimality and analysis of block, Latin square and change-over designs when the standard assumption of independence of observations is likely to be violated. The approach is to postulate a correlation model for the observations and then apply generalized least squares for estimating treatment effects. The optimality criterion is a functional of the variance-covariance matrix of the treatment effect estimates. This approach and other related techniques, such as Papadakis' nearest neighbour method, Wilkinson's moving blocks method, etc., have been found to be powerful tools for increasing the precision of treatment comparisons in actual field experiments.

Similar situations arise in the case of repeated measures data or cross-over experiments data where an organism is being observed repeatedly over the time. The temporal nature of measurements calls for accommodation of possible correlation in the design and analysis of such studies.

Selected Publications

Gill, P.S. and Shukla, G.K. (1985). "Efficiency of nearest neighbour balanced block designs for correlated observations." *Biometrika*, **72**, 539-544.

Gill, P.S. and Shukla, G.K. (1987). "Optimal change-over designs for correlated observations." *Communications in Statistics - Theory and Methods*, **16**, 2243-2261.

Gill, P.S. (1990). "A Monte Carlo simulation study of analyses of block designs under a correlated errors model." *Communications in Statistics - Simulation and Computation*, **19**, 175-189.

Gill, P.S. (1992). "A note on modelling the covariance structure of repeated measurements." *Biometrics*, **48**, 965-968.

Gill, P.S. (1992). "Balanced change-over designs for autocorrelated observations." *Australian Journal of Statistics*, **34**, 415-420.

Nhu D. Le

HONORARY RESEARCH ASSOCIATE

B.Sc. (1984), The University of British Columbia

M.Sc. (1986), The University of British Columbia

Ph.D. (1990), University of Washington

Post-Doctoral Fellow, University of British Columbia (1990-1992)

Biostatistician, British Columbia Transplant Society (1991-1992)

Research Scientist, British Columbia Cancer Agency (1992-present)

The University of British Columbia (1992-present)

Research Interests

My current research interests are primarily in the areas of spatial statistics and time series with emphasis on applications of statistics to occupational and environmental health problems. The work motivated by occupational oncology studies carried out at the B.C. Cancer Research Centre, concerns the possibility of characterizing occupational cancer risk factors and identifying potential carcinogens in the work place relevant to the specific industrial context of B.C. The overall objective is the reduction of risk. Issues arising from environmental studies led to my participation in the development of a theory for spatial interpolation. The work allows for the derivation of spatial prediction distributions for regions with no monitoring station using data available at only a few monitoring stations. This theory is particularly useful in environmental health applications where impact assessment of pollution on public health is needed and pollution data are available only at a few locations. Another consequence of this work is the development of methods for redesigning environmental monitoring network where one may want to either extend or contract an existing network. Various theoretical extensions of the theory and their potential applications are being explored. Other research topics of interest include the use of mixture distributions in modelling overdisperse Poisson counts or non-Gaussian time series, and the use of robust Bayes factors in model comparison for autoregressive processes.

Selected Publications

Le, N.D. and Zidek, J.V. (1992). "Interpolation with uncertain spatial covariances: A Bayesian alternative to Kriging." *Journal of Multivariate Analysis,* **43**, 351-374.

Le, N.D., Leroux, B.G. and Puterman, M.L. (1992). "Exact likelihood evaluation in a Markov mixture model for time series of seizure counts." *Biometrics,* **48**, 317-323.

Montaner, J.S.G., Le, N.T., Le, N.D., Craib, K.J.P. and Schechter, T.M. (1992). "Application of the World Health Organization system for HIV infection in a cohort of homosexual men in developing a prognostically meaningful staging system." *AIDS,* **6**, 719-724.

Le, N.D. and Zidek, J.V. (1992). "Network designs for monitoring multivariate random spatial fields." In: *Proceeding of the 4th International Meeting of The Basque Country,* J.P. Vilaplena (ed.). To appear.

Le, N.D., Martin, R.D. and Raftery, A.E. (1993). "Robust model comparison for autoregressive processes with robust Bayes factors." Technical Report 123, Department of Statistics, UBC.
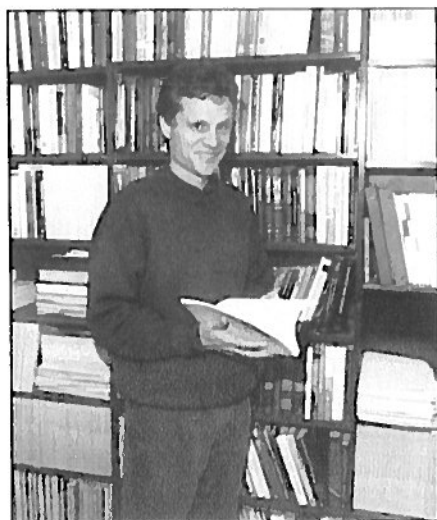
Piet de Jong

ASSOCIATE MEMBER

B.Ec. (1971), University of Sydney

Ph.D. (1985), LaTrobe University, Melbourne

Lecturer, University of Sydney (1975-77)

Assistant Professor, Simon Fraser University (1977-79)

The University of British Columbia (1979-present)

Research Interests

Is there global warming? Do seat belts save lives? Will more income taxation increase taxation revenue? Will mortgage rates decline? How many hospital beds will we need for AIDS patients in the year 2000? Are the present funding formulas for the Canada Pension Plan sufficient to provide required future benefits? Is such and such a spacecraft (or missile) on target?

Questions of this kind are often answered by analyzing relevant time series of data: yearly temperatures, road fatality statistics and seat belt usage statistics over time, monthly interest rates, etc. When analyzing time series we need an appropriate statistical model. My primary research interest is on a certain class of time series models - namely models that are linear or linear after transformation or conditioning.

The linear class of time series models is both practical and widely applicable. An appropriate linear model can often come to grips with questions such as the above. Much is known about certain linear time series models. However, my interest focuses on what is generally true about these models: How they may be fitted or how they can be used for prediction and inference.

Two crucial algorithms for linear time series models are the Kalman filter and the smoothing filter. These two algorithms sequentially process time series data allowing for the fitting and evaluation of models. I am interested in theoretical properties of these algorithms, their actual implementation details as well as how they perform in practice. The relevance of these algorithms is increasing as cheaper computing power continually opens up further areas of potential application. At the same time the deeper understanding of these two algorithms reveals further scope for application.

Apart from the theory and implementation of linear time series models, I am also interested in the application of the models to particular subject areas such as actuarial science, ecology, finance, and weather and ocean patterns. Other areas of interest are pairing models, Markov chains and Forensic statistics.

Selected Publications

de Jong, P., Greig, M. and Madan, D.B. (1983). "Testing for random pairing." *Journal of the American Statistical Association*, **78**, 332-336.

de Jong, P. and Greig, M. (1984). "First order Markov chains with a zero diagonal transition matrix." *Biometrics*, **40**, 101-108.

de Jong, P. (1988). "The likelihood for a state space model." *Biometrika*, **75**, 165-169.

de Jong, P. (1989). "Smoothing and interpolation with the state space model." *Journal of the American Statistical Association*, **84**, 1085-1088.

de Jong, P. (1991). "The diffuse Kalman filter." *Annals of Statistics*, **2**, 1073-1083.

Priscilla E. Greenwood

ASSOCIATE MEMBER

B.A. (1959), Duke University

M.A. (1961), University of Wisconsin

Ph.D. (1963), University of Wisconsin

Assistant Professor, University of Wisconsin (1963-64)

Associate Professor, North Carolina College (1964-66)

The University of British Columbia (1966-present)

Research Interests

Some topics of my past research in stochastic processes are variations of processes, optional stopping, Wiener-Hopf factorization, domains of attraction, Markov splitting, harmonic renewal measures, excursions of processes, set-indexed Brownian motion, weak convergence of set-indexed processes and of evolving random fields. More recently some of my research is in statistics of stochastic processes. Some topics I have worked on are contiguity and convergence of experiments, asymptotic efficiency in the Hajek-Le Cam sense, partial likelihood, efficient estimating equations, optimal estimation for stochastic process families with critical points, misspecified models, empirical estimators for Markov and related processes.

Current research projects include asymptotically efficient estimation in a locally uniform Bahadur (large deviations) sense, efficient estimation from Monte Carlo-Markov chain (e.g., Gibbs' sampler) data, model selection for stochastic process families using efficient estimators under model misspecification. Another direction of statistical research is in statistical physics, where random field measures, e.g., Gibbs' measures, can be interpreted in terms of likelihoods. It should be possible to exploit this structure together with likelihood methods for stochastic processes to investigate efficient inferences for evolving random fields and to study the question of inference near critical points where a physical change-of-state occurs.

Selected Publications

Greenwood, P.E. and Shiryaev, A.N. *Contiguity and the Statistical Invariance Principle.* Gordon and Breach, 1985.

Greenwood, P.E. and Ossiander, M. (1991) "Functional Convergence of Evolving Random Fields." *IMS Lecture Notes Series* **18**, 66-99.

Greenwood, P.E. and Wefelmeyer, W. (1993). "Asymptotic minimax results for stochastic process families with critical points." *Stochastic Processes and their Applications*, **44**, 107-116.

Greenwood, P.E. and Wefelmeyer, W. "Efficiency of empirical estimators for Markov chains." *Annals of Statistics.* To appear.
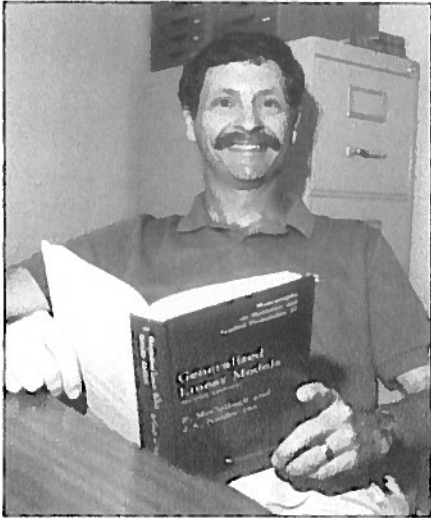
Research Interests

My current research interests are in time series econometrics and non-linear time series. Specifically, I am interested in testing for unit roots in the context of non-stationary series; that is, are series difference stationary? Other interests are in developing non-linear models for econometric data.

Selected Publications

McCabe, B.P.M. (1990). "An extension of Anderson's multiple decision procedure using invariance." *Probability and Statistics Letters*, **9**, 119-124.

McCabe, B.P.M. and Leybourne, S.J. (1993). "Testing for parameter variation in non-linear regression models." *Journal of the Royal Statistical Society*, **B55**, 133-164.

Brendan McCabe

ASSOCIATE MEMBER

B.A. (1974), Trinity College Dublin

M.Sc. (1975), London School of Economics

Ph.D. (1989), University of Amsterdam

Lecturer, University of Leeds, 1976-1988

Senior Lecturer, University of Sydney, 1988-1989

The University of British Columbia, (1989-present)

Martin L. Puterman

ASSOCIATE MEMBER

A.B. (1966), Cornell University

Ph.D. (1971), Stanford University

Visiting Assistant Professor and Assistant Professor, University of Massachusetts (1971-73)

Visiting Assistant Professor, The University of British Columbia (1974-75)

The University of British Columbia (1975-present)

Research Interests

My primary research areas are applied biostatistics and stochastic dynamic programming.

My statistical research is motivated by medical applications, especially studies in which each subject's status can be described by a sequence of discrete outcomes. Examples include infectious disease clinical trials in which a subject is observed to be infected or not at successive visits and studies of epileptic subjects in which data consists of daily seizure counts on subjects.

My research in this area focuses on discrete mixture models. Discrete mixture models provide plausible models for data in which the variation exceeds that predicted by binomial and Poisson models. My current focus is on independent and correlated Poisson mixture models which include covariate information.

My research on dynamic programming focuses on applications, algorithms and computation. I have recently completed the book "Markov Decision Processes" which provides a comprehensive study of theory, application and computation for these models.

Selected Publications

Leroux, B.G. and Puterman, M.L. (1992). "Maximum penalized likelihood estimation for independent and Markov dependent mixture models." *Biometrics,* **48**, 545-558.

Le, N.D., Leroux, B.G. and Puterman, M.L. (1992). "Exact likelihood evaluation in a Markov mixture model for time series of seizure counts." *Biometrics,* **48**, 317-323.

Puterman, M.L. (1993). *Markov Decision Processes.* John Wiley and Sons.

Gribble, M.J. and Puterman, M.L. (1993). "Prohylaxis of urinary tract infection in persons with recent spinal cord injury: A prospective, randomized, double-blind, placebo-controlled study of trimethoprim-fulfamethoxazole." *The American Journal of Medicine.* To appear.

M.Sc. Theses

YANLING ZUO (1990)

SUPERVISOR: N.E. Heckman

TITLE: Monotone Regression Functions

ABSTRACT: In some applications, we require a monotone estimate of a regression function. In others, we want to test whether the regression function is monotone. To solve these problems, the regression function is estimated in one of two ways: via least-squares and via penalized likelihood. Both estimates require the choice of a "smoothing parameter", which is chosen via modifications of Akaike Information Criteria (for least-squares) and Generalized Cross Validation (for penalized likelihood). These estimates are used in testing for monotonocity. Simple properties of the estimates are studied and simulations are carried out to compare the estimates. The techniques are applied to two data sets.

ZDENEK PATAK (1990)

SUPERVISOR: R.H. Zamar

TITLE: Robust Principal Component Analysis via Projection Pursuit

ABSTRACT: In principal component analysis (PCA), the principal components (PC) are linear combinations of the variables that minimize some objective function. In the classical setup the objective function is the variance of the PC's, which can be easily upset by outlying observation; we propose a different objective function, yielding MM- and t- estimates, which have both high breakdown point and efficiency.

As a good initial estimate in the minimization procedure, we propose an orthogonal version of the least median of squares (Rousseeuw and Leroy, 1987) and a new methods that is orthogonal equivariant, robust and easy to compute. Orthogonal regression and detection of multivariate outliers are discussed as possible applications of PCA.

AL-KARIM LADAK (1990)

SUPERVISORS: M. Delampady
                & A.J. Petkau

TITLE: Resampling-Based Variance Estimators in Ratio Estimation with Application to Weigh Scaling

ABSTRACT: Weigh scaling is a method of estimating the total volume of timber harvested from a given region. Jackknife- and bootstrap-based estimators of the variance of the estimated total volume are considered. Weighted versions are derived using influence function and Fisher Information matrices. Empirical studies of these variance estimators, with particular emphasis on small sample properties and robustness with respect to both the homoscedastic variance and zero-intercept population characteristics, reveal that these estimators have small negative biases for small sample sizes and are robust with respect to heteroscedasticity.

YOUSHI CHEN (1991)

SUPERVISOR: H. Joe

TITLE: Estimation with Multivariate Extreme Value Distributions, with Applications to Environmental Data

ABSTRACT: Several parametric families of multivariate extreme value distributions (Hüsler and Reiss 1989, Tawn 1990, Joe 1990a, 1990b) have been proposed recently. Applications to multivariate extreme value data sets are needed to assess the adequacy of the known families in their fit to data. Different families are compared in their range of multivariate dependence and their ease of use for maximum likelihood estimation. Some useful conclusions have been made from experience with several environmental data sets.

RUI FANG (1991)

SUPERVISOR: A. Coldman

TITLE: Simultaneous and Sequential ROC Analysis for Multiple Diagnostic Tests

ABSTRACT: Relative or receiver operating characteristic (ROC) analysis is a simple procedure which can be used to measure the accuracy of diagnostic tests. Diagnostic tests are often used to classify an individual as belonging to one of two populations. Based on statistical decision theory, ROC was first developed to evaluate the performance of electronic signal detection, and has been used to evaluate the accuracy of diagnostic tests. The ROC theory for evaluating one single test, or comparing individual tests is reasonably well understood. The question arises in cases where multiple tests are available as to whether some combination of the tests are better than any single one. In this paper, two ROC procedures of evaluating the aggregate performance of multiple diagnostic tests were presented, one is for evaluating simultaneous multiple diagnostic tests, and the other is for sequential diagnostic tests. These procedures are illustrated using a breast cancer data set.

JINKO GRAHAM (1992)

SUPERVISOR: A. J. Petkau

TITLE: Longitudinal Analysis for Binary and Count Data

ABSTRACT: Longitudinal data sets consist of repeated observations of an outcome over time, and a corresponding set of covariates for each of many subjects. Appropriate analyses must take into account the correlation in the repeated observations, but this structure may not be well understood, making parametric modelling difficult.

Two methods which require only minimal assumptions about the true correlation structure and allow the inclusion of time-dependent covariates based on the use of a "working" likelihood

and on quasi-likelihood theory. The latter method can be applied to any longitudinal response with univariate marginal distributions for which the quasi-likelihood formulation is sensible. As is illustrated using the results from an experiment on hummingbird learning, these methods enable much more information to be extracted than more traditional analysis-of-variance methods, and therefore provide useful and powerful tools for researchers in this subject area.

PAUL GUSTAFSON (1992)

SUPERVISOR: M. Delampady

THESIS TITLE: Hierarchical Modelling of Multivariate Survival Data

ABSTRACT: Hierarchical models based on conditional independence are investigated as a means of modelling multivariate survival times. The model structure follows Clayton (1978), Hougaard (1986b), and Oakes (1986, 1989). Both approximate Bayesian and maximum likelihood estimation in these models is investigated via simulation. Predicting a component of a response vector on the basis of other components of the response and other vectors is also studied, using Bayes and empirical Bayes methods. Application to real data is detailed, using the techniques of estimation and prediction discussed.

SONIA MAZZI (1992)

SUPERVISOR: R.H. Zamar

TITLE: A New Measure of Quantitative Robustness

ABSTRACT: The Gross-Error Sensitivity (GES) and the Breakdown Point (BP) are two measures of quantitative robustness which have played a key role in the development of the theory of robustness. Both can be derived from the maximum bias function B(e) and constitute a two-number summary of this function.

A new robustness quantifier, the breakdown rate (BR), that summarizes the behavior of B(e) for e near BP will be introduced. The BR for several families of robust estimates of regression will be presented and the increased usefulness of the three-number summary (GES, BP, BR) for comparing robust estimates will be illustrated by several examples.

SAULATI KOMUNGOMA (1992)

SUPERVISORS: N.E. Heckman
                & J.V. Zidek

TITLE: Assessment of the Quality for the NADP/NTN Data Based on their Predictability

ABSTRACT: Three methods for predicting the ion concentrations at a particular station from those at other stations in a network are assessed using data from a major environmental monitoring network. Stations were ranked according to the degree of predictability of their concentration levels. This ranking would constitute an ordering if, hypothetically, a termination of some stations became necessary.

Our study used monthly volume weighted mean concentrations for each of the three selected ions, investigated one at a time. The large number of stations involved and paucity of data forced a cluster-by-cluster analysis of the stations. For these analyses missing data were imputed using a regression-based imputation strategy.

BRAD McNENEY (1992)

SUPERVISOR: A.J. Petkau

TITLE: Overdispersion in Poisson Regression

ABSTRACT: Previous investigations of possible relationships between air quality and human health in the community of Prince George, British Columbia have been based on hospital admissions and emergency room visits for respiratory problems during the period April 1984 to March 1986. The study described here was a follow-up based on emergency room visits data for the period from April 1986 to March 1988.

The main part of the analyses involved the use of Poisson regression models with an extension to account for over-dispersion in the data. An extension simulation study was subsequently carried out to assess the methods used in these analysis and to compare these to other possible estimators and test statistics that can be employed in the analysis of over-dispersed Poisson data.

ED SUSKO (1992)

SUPERVISOR: J. Liu

TITLE: Segmented Regression Modelling with an Application to German Exchange Rate Data

ABSTRACT: Segmented regression models are the topic of this thesis. These are regression models in which the mean response is thought to be linear in the explanatory variables within regions of a particular explanatory variable. A criterion for estimating the number of segments in a segmented model is given and the consistency of this estimator is established under rather general conditions.

Segmented models are fitted to German exchange rate data and forecasting results obtained from these models are compared with forecasting results from widely used models in exchange rate prediction. The segmented models tend to perform better than models that have been established in the literature, notably, the random walk model.

RICK WHITE (1992)

SUPERVISOR: R.H. Zamar

TITLE: The Detection and Testing of Multivariate Outliers

ABSTRACT: To detect and test for the presence of multivariate outliers, a sequential multivariate scale-ratio test is proposed. It is based on the "ratio" of a non-robust estimate and a robust estimate

of scatter and is applied in a forward fashion, removing the most extreme point at each step, until the test fails to indicate the presence of outliers. The thesis also shows that this procedure has level a when applied to an uncontaminated sample, is uneffected by swamping or masking and is accurate in detecting outliers.

Ph.D. Theses

ANDREW J. COLDMAN (1987)

SUPERVISOR: A.J. Petkau

TITLE: The Development of Resistance To Anticancer Agents.

ABSTRACT: This thesis explores the mechanism of resistance of tumour cells to chemotherapeutic agents using a discrete time Markov branching process which models the growth of the tumour and resistant cells.

Probability generating functions are derived which describe the behaviour of the process after an arbitrary sequence of drug treatments. Expressions for the probability of ultimate extinction of the stem cell compartment are derived for a number of experimental situations and the effects of variation in the parameter values are examined. This work is extended to the case where two anticancer drugs are available. These models are applied to data collected on the L1210 leukemia treated by the drugs Cyclophosphamide and Arabinosylcytosine. In both cases the data provides evidence that resistant cells arise spontaneously with a frequency of approximately 10^{-7} per division.

BRIAN LEROUX (1989)

SUPERVISOR: M. Puterman

TITLE: Maximum Likelihood Estimation for Mixture Distributions and Hidden Markov Models

ABSTRACT: This thesis deals with computational and theoretical aspects of maximum likelihood estimation for data from a mixture model and a hidden Markov model.

In the mixture model, a maximum penalized likelihood methods is proposed and shown to produce a consistent estimator of the unknown mixing distribution. In the hidden Markov model, a sequence of states is modelled by a Markov chain and the observed variables are conditionally independent with distributions determined by the states. The use of the EM algorithm for finding local maxima of the likelihood function is described. The consistency of maximum likelihood estimators is proved, along with preliminary results on identifiability, ergodicity, entropy convergence and corss entropy convergence.

SHIYING WU (1992)

SUPERVISOR: J. Liu

THESIS TITLE: Asymptotic Inference for Segmented Regression Models

ABSTRACT: This thesis deals with the estimation of segmented multivariate regression models. A segmented regression model is a regression model which has different analytical forms in different regions of the domain of the independent variables. Without knowing the number of these regions and their boundaries, we first estimate the number of these regions by using a modified Schwarz' criterion. Under fairly general conditions, the estimated number of regions is shown to be weakly consistent. We then estimate the change points or "thresholds". It is shown that the estimates of the thresholds converge at the rate of $Op(ln^2 n/n)$, if the model is discontinuous at the thresholds, and $Op(n^{-1/2})$ if the model is continuous. In both cases, the estimated regression coefficients and residual variances are shown to be asymptotically normal. As an illustration, a segmented bivariate regression model is fitted to real data and the relevance of the asymptotic results is examined through simulation studies. The identifiability of the segmentation variable is also discussed. Under different conditions, two consistent estimation procedures of the segmentation variable are given. The results are then generalized to the case where the noises are heteroscedastic and autocorrelated.